

CT 사진으로 암 진단하기

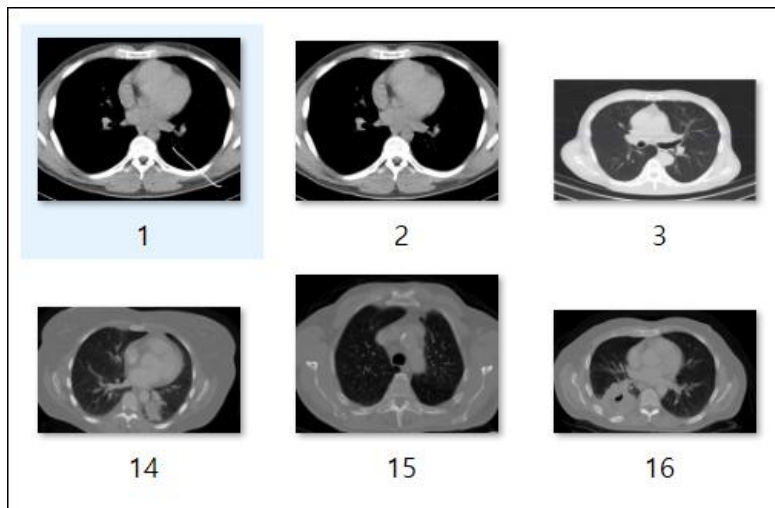


소프트웨어학부
진혜진

■ Classification

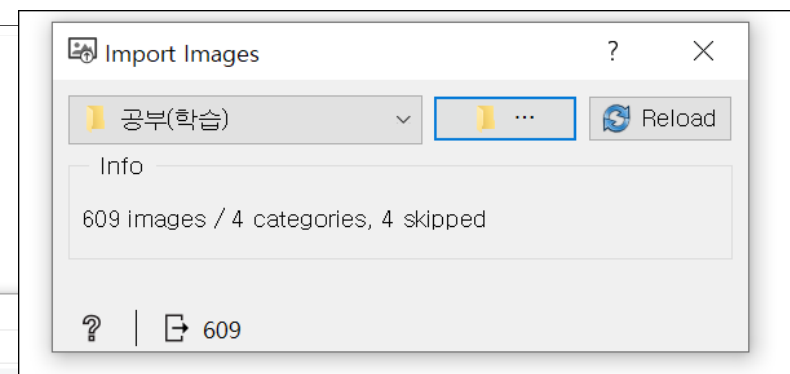
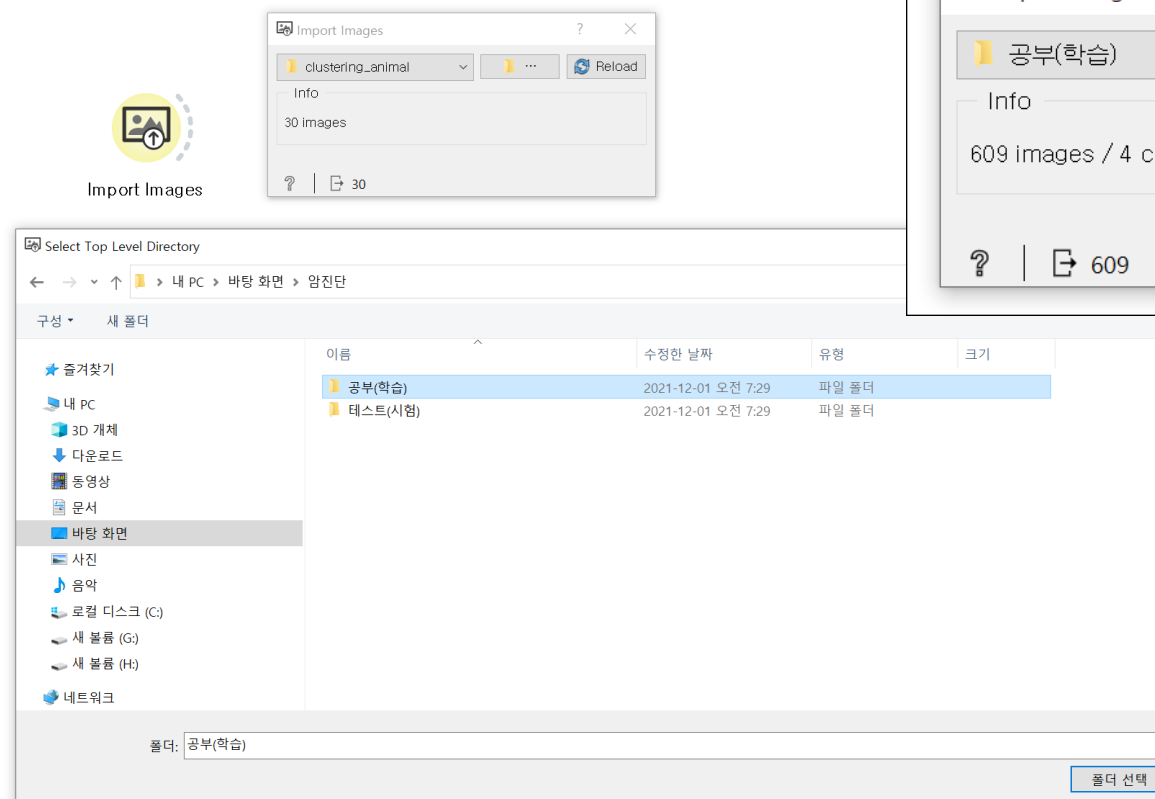
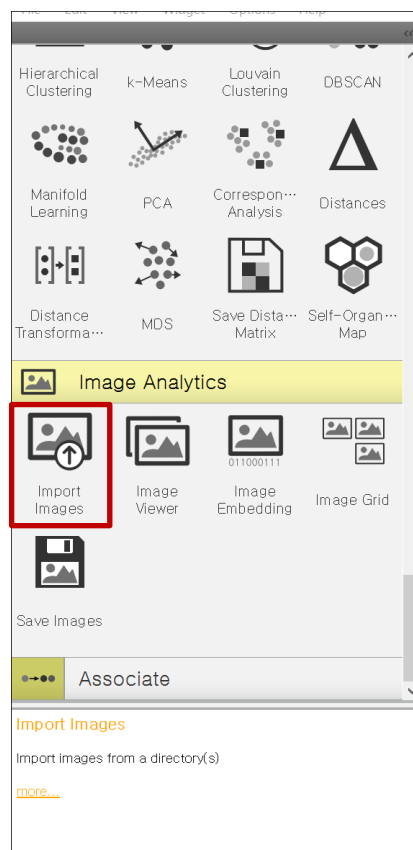
- 보기가 주어져 있는 객관식 문제를 푸는 것과 비슷하다.
- 보기가 정해져 있는 예측들은 모두 분류이다.
- 데이터를 특정 보기(범주)로 분류하는 것으로, 새로운 데이터가 어떤 범주에 속할지 예측할 수 있는 기법이다.
- 경영활동에서의 의사결정 뿐만 아니라 의학, 법률 등 여러 분야에 적용 가능하다.
 - 회사에서 이직할 직원을 미리 예측해볼 수도 있고, CT 사진을 통해 암 질환 여부 나아가 암 질환 종류까지도 확인할 수 있다.

- CT 사진 중에는 암 환자도 있고 아닌 사람도 있다.
- 이런 사진을 구분 할 때도 인공지능을 사용할 수 있다.
- 인공지능에게 CT사진(문제)들을 보여주고, 각 사진 별로 암 질환 여부 및 종류(정답)을 알려주면서 학습(공부)시키는 것이다.
- 인공지능은 의학에 대한 지식은 없지만, 그냥 그림만으로도 암인지 아닌지, 암이라면 어떤 암인지 등을 예측하기 시작한다.

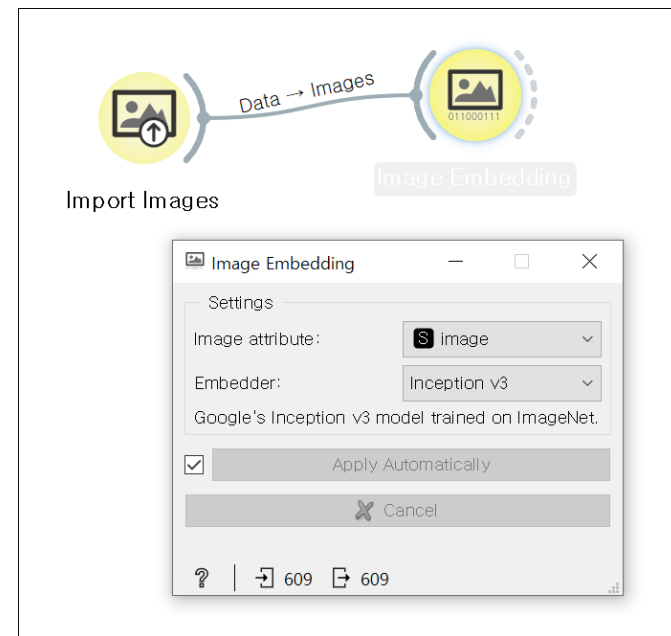
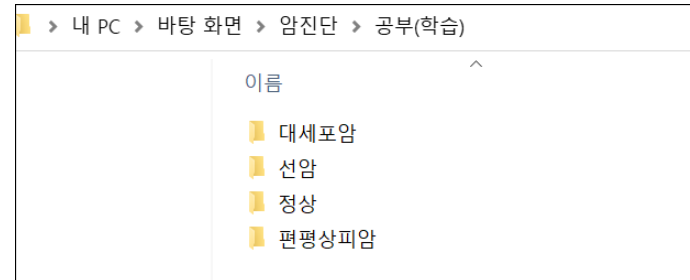


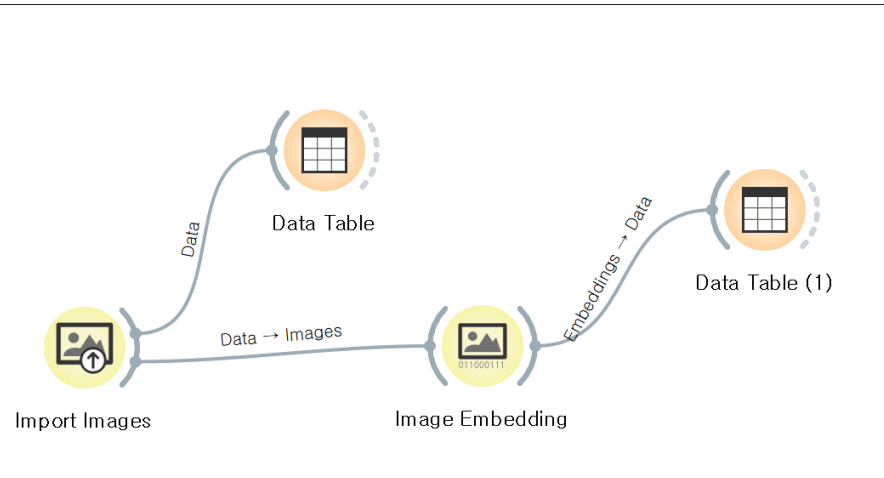
■ 데이터 셋 다운로드

- 암 진단 압축파일 다운받아서 바탕화면에 풀기
- Image Analytics 메뉴에서 Import Image 선택
- 학습에 필요한 데이터(CT사진)를 불러온다. 일단 학습을 시켜야 하므로, 공부(학습)을 선택한다.



- 객관식 예측 분류 분석의 경우, 인공지능을 공부시킬 때 이름표(정답)을 달아줘야 하는데, 오렌지에서는 이미지 분류의 경우 폴더 명을 이름표로 인식한다.
- 정답 이름으로 폴더를 만들고, 그 안에 해당 이미지를 집어넣으면, 그 이미지들은 모두 그 이름표가 붙은 것으로 알아듣는다.
- 임베딩
 - 이미지는 바로 읽을 수가 없어서, 컴퓨터가 이해할 수 있는 형태로 바꿔줄 필요가 있다.
- 구글의 Inception v3라는 우수한 모델이 기본으로 설정되어 있다.





Data Table

Info
609 instances (no missing data)
No features
Target with 4 values
5 meta attributes

Variables
☒ Show variable labels (if present)
☐ Visualize numeric values
☒ Color by instance classes

Selection
☒ Select full rows

origin	category	image name	image sDrive/바탕 화면/ image	size	width	height
1	대세포암	000002	대세포암w000...	72317	423	213
2	대세포암	000003 (3)	대세포암w000...	80313	437	258
3	대세포암	000003 (4)	대세포암w000...	81264	443	260
4	대세포암	000009 (3)	대세포암w000...	83258	440	270
5	대세포암	000009 (4)	대세포암w000...	81078	423	267
6	대세포암	000010	대세포암w000...	90327	369	281
7	대세포암	000015	대세포암w000...	81727	401	266
8	대세포암	000016 (3)	대세포암w000...	89509	414	251
9	대세포암	000016 (4)	대세포암w000...	96450	412	279
10	대세포암	000017	대세포암w000...	93307	394	266
11	대세포암	000018 (2)	대세포암w000...	60213	402	189
12	대세포암	000019 (3)	대세포암w000...	88691	439	258
13	대세포암	000019 (4)	대세포암w000...	84443	444	241
14	대세포암	000020 (3)	대세포암w000...	84937	426	259
15	대세포암	000020 (4)	대세포암w000...	91046	412	297
16	대세포암	000021	대세포암w000...	85076	385	265
17	대세포암	000023 (2)	대세포암w000...	59743	432	174

Data Table (1)

Info
609 instances (no missing data)
2048 features
Target with 4 values
5 meta attributes

Variables
☒ Show variable labels (if present)
☐ Visualize numeric values
☒ Color by instance classes

Selection
☒ Select full rows

hidden origin	category	image name	image sDrive/바탕 화면/ image	size	width	height	n0 True	n1 True	n2 True	n3 True
1	대세포암	000002	대세포암w000...	72317	423	213	0.29367	0.0276482	0.693917	1.08905
2	대세포암	000003 (3)	대세포암w000...	80313	437	258	0.247316	0.0117799	0.388568	0.348236
3	대세포암	000003 (4)	대세포암w000...	81264	443	260	0.146152	0.0174174	0.406528	0.359734
4	대세포암	000009 (3)	대세포암w000...	83258	440	270	0.135613	0.059938	0.657129	0.243374
5	대세포암	000009 (4)	대세포암w000...	81078	423	267	0.219178	0.00819431	0.686168	0.22643
6	대세포암	000010	대세포암w000...	90327	369	281	0.497037	0.0545665	0.706522	0.338154
7	대세포암	000015	대세포암w000...	81727	401	266	0.305263	0.0100914	0.551818	0.357818
8	대세포암	000016 (3)	대세포암w000...	89509	414	251	0.264506	0.239051	0.962501	0.0152599
9	대세포암	000016 (4)	대세포암w000...	96450	412	279	0.32365	0.155225	1.09032	0.00726441
10	대세포암	000017	대세포암w000...	93307	394	266	0.2861	0.145969	0.646866	0.0974206
11	대세포암	000018 (2)	대세포암w000...	60213	402	189	0.050852	0.11623	0.668737	0.21378
12	대세포암	000019 (3)	대세포암w000...	88691	439	258	0.10539	0.0123145	0.523354	0.208393
13	대세포암	000019 (4)	대세포암w000...	84443	444	241	0.0835931	0.000177786	0.561296	0.29155
14	대세포암	000020 (3)	대세포암w000...	84937	426	259	0.227674	0.0373076	0.526607	0.0390441
15	대세포암	000020 (4)	대세포암w000...	91046	412	297	0.227413	0.116501	0.393722	0.0827895
16	대세포암	000021	대세포암w000...	85076	385	265	0.460819	0.145445	0.439538	0.350832
17	대세포암	000023 (2)	대세포암w000...	59743	432	174	0.176704	0.039811	0.278675	0.295996
18	대세포암	000024 (2)	대세포암w000...	50944	453	186	0.319239	0	1.31382	0.245249
19	대세포암	000026 (2)	대세포암w000...	92055	417	256	0.351135	0.18738	0.450212	0.0983565
20	대세포암	000026	대세포암w000...	89289	382	291	0.519009	0.152752	0.598074	0.354054
21	대세포암	000027 (2)	대세포암w000...	69893	352	227	0.137208	0.232939	0.723243	0.0808237
22	대세포암	000027	대세포암w000...	75786	381	231	0.541321	0.133514	0.511328	0.421374
23	대세포암	000031 (2)	대세포암w000...	65905	403	207	0.178222	0.052377	0.752682	0.352172
24	대세포암	000031	대세포암w000...	87421	361	277	0.608212	0.10382	0.915126	0.0922305

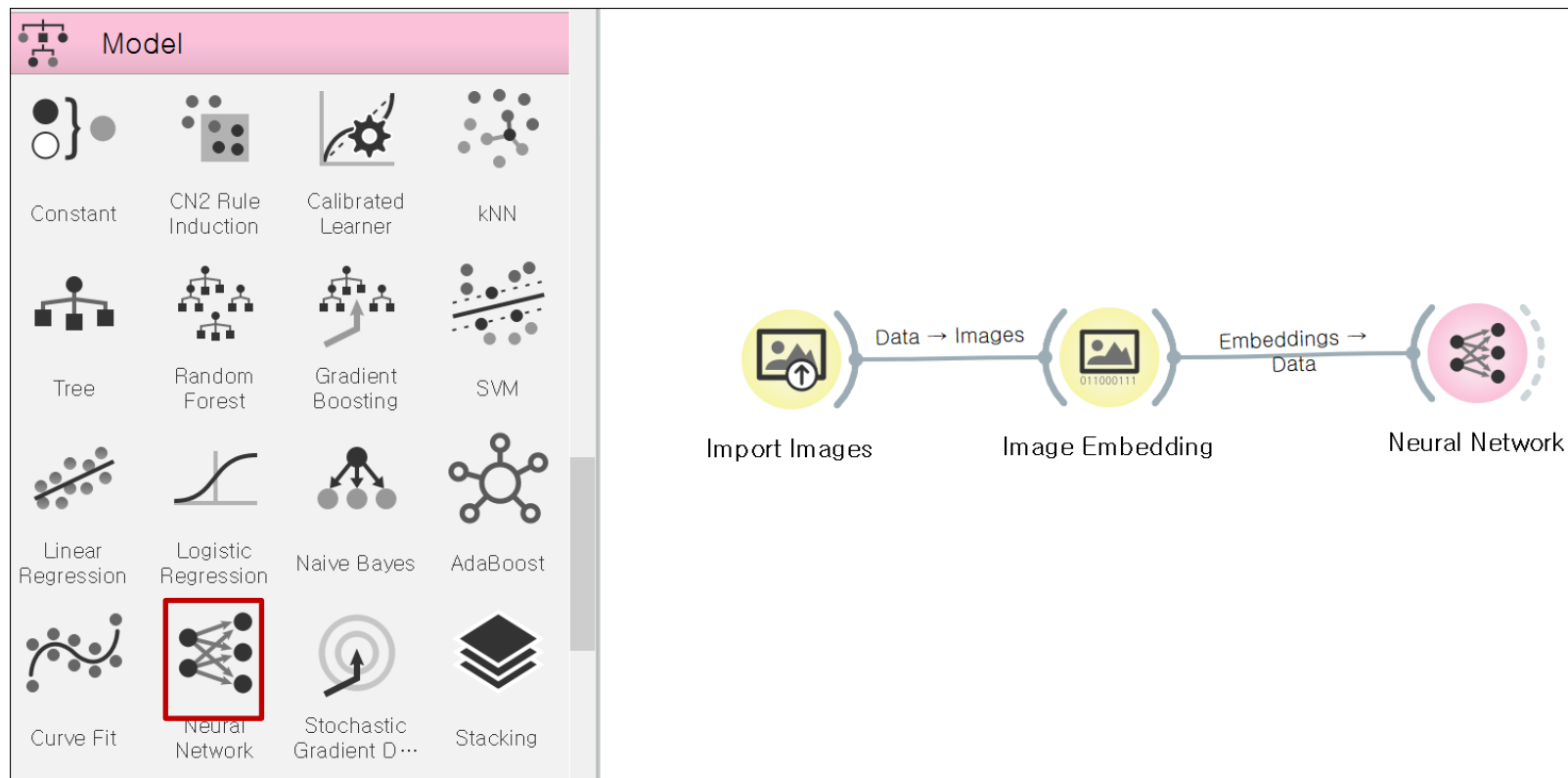
Restore Original Order

☒ Send Automatically

609 1 | 609

50944	453	186
92055	417	256
89289	382	291
69893	352	227
75786	381	231
65905	403	207
87421	361	277
96060	377	317

■ 모델 사용하기



■ Test and Score

- 인공지능 모델들이 열심히 계산하고 공부한다. 계산이 끝나면 Test and Score를 더블 클릭해 보자.

■ 학습 성능

Evaluate

Test and Score

Predictions

Confusion Matrix

ROC Analysis

Lift Curve

Calibration Plot

Unsupervised

Image Analytics

Associate

Test and Score

Cross-validation accuracy estimation.

[more...](#)

Test and Score

Sampling

☒ Cross validation

Number of folds: 5

☒ Stratified

☐ Cross validation by feature

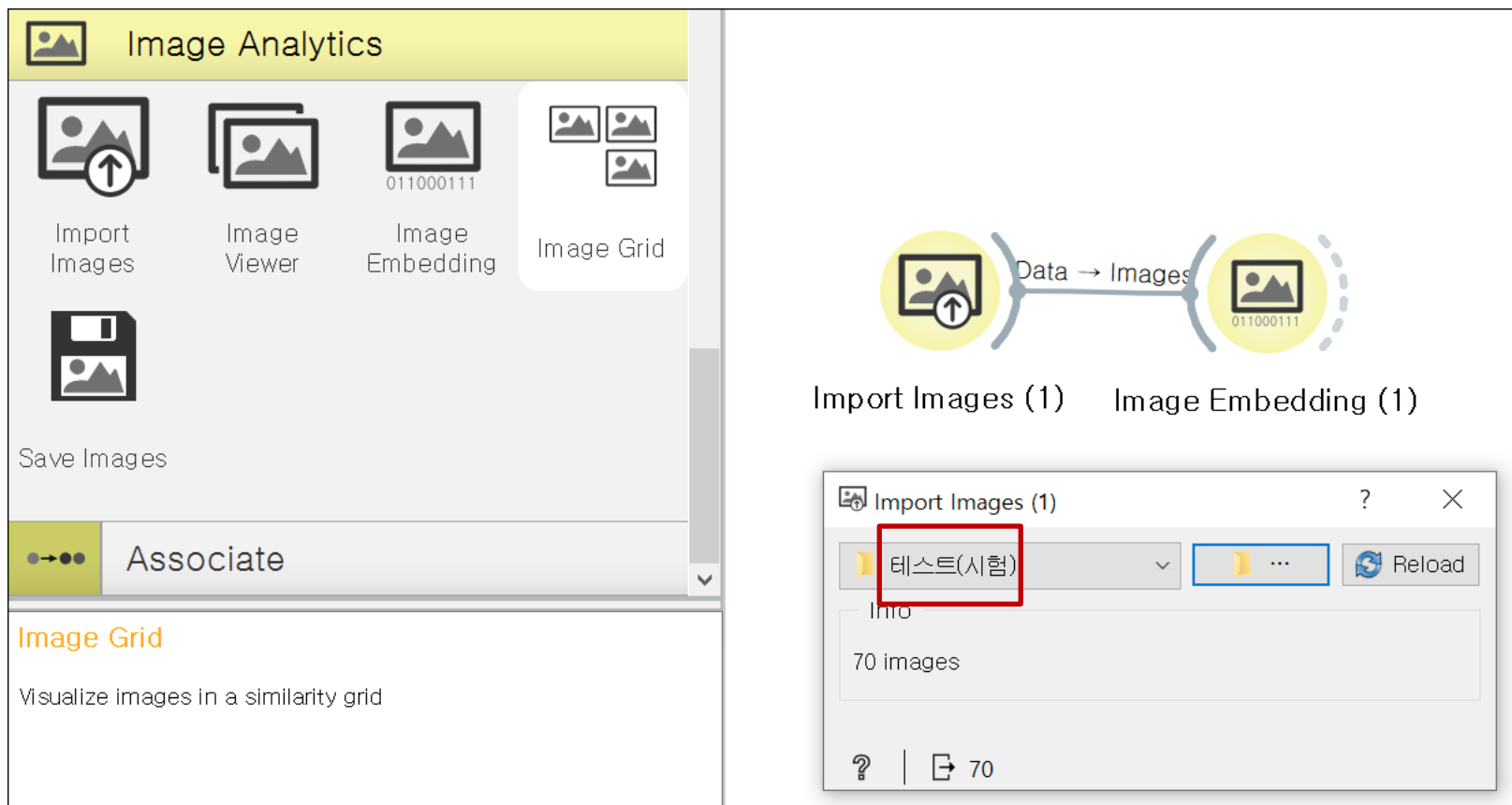
☐ Random sampling

Repeat train/test: 10

Evaluation Results

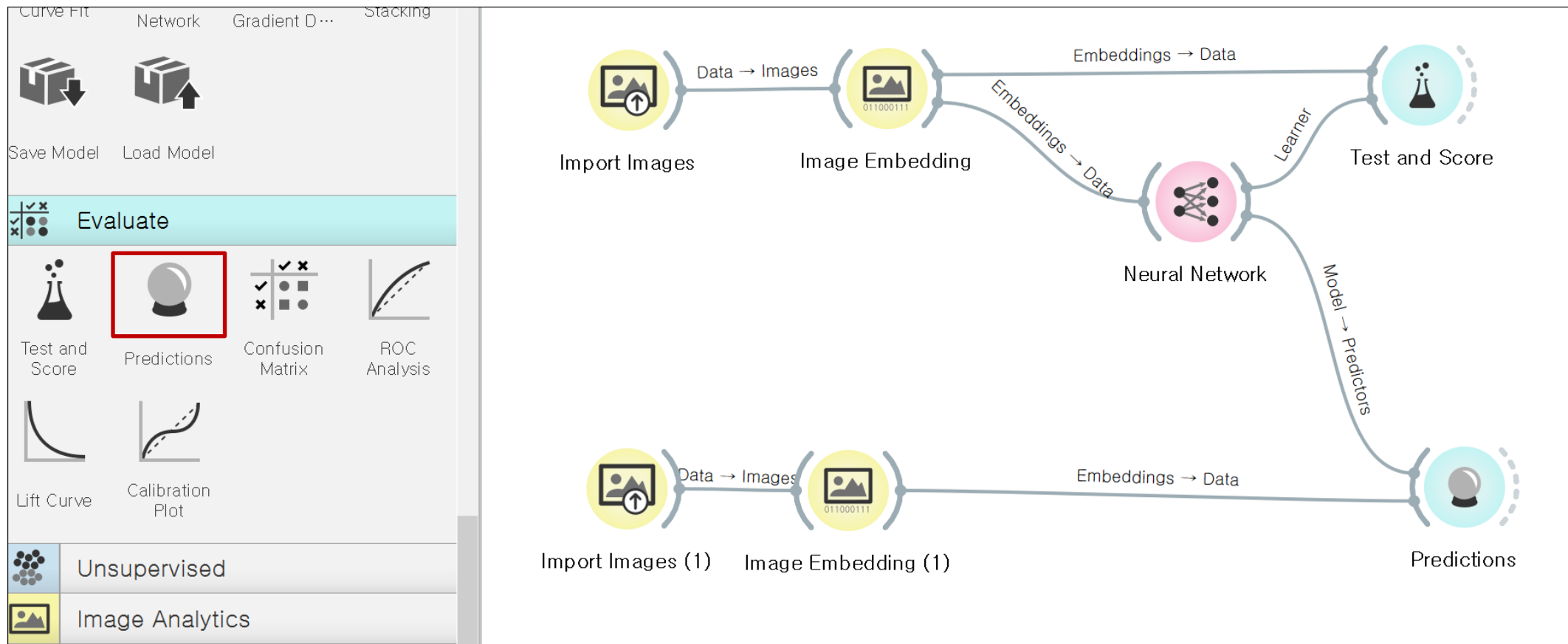
Model	AUC	CA	F1	Precision	Recall
Neural Network	0.974	0.878	0.878	0.879	0.878

■ 인공지능 모델을 학습시켰으니 실전 예측을 해보자.



The screenshot displays the 'Image Analytics' application interface. On the left, a sidebar contains icons for 'Import Images', 'Image Viewer', 'Image Embedding' (labeled '011000111'), 'Image Grid', and 'Save Images'. Below these is an 'Associate' button. The main workspace shows a workflow diagram with two nodes: 'Import Images (1)' and 'Image Embedding (1)', connected by a 'Data → Images' arrow. A dialog box titled 'Import Images (1)' is open, showing a dropdown menu with '테스트(시험)' (Test) selected, a 'Reload' button, and a list of '70 images'.

- Predictions을 불러와 변환된 데이터와 앞서 공부시킨 모델로 연결시킨다.
- 계산이 끝나면 Predictions을 더블 클릭해 예측 값을 확인할 수 있다.



Predictions

Show probabilities for

대세포암
선암
정상
편평상피암

	Neural Network
1	0.00 → 정상
2	0.00 → 정상
3	0.11 → 대세포암
4	0.08 → 편평상...
5	0.99 → 선암
6	0.00 → 편평상...
7	0.22 → 대세포암
8	1.00 → 선암
9	0.95 → 선암
10	0.11 → 편평상...
11	0.00 → 편평상...
12	0.00 → 정상

image name	image
1	1.png
10	10.png
11	11.png
12	12.png
13	13.png

Predictions

Show probabilities for

대세포암
선암
정상
편평상피암

	Neural Network
1	0.00 → 정상
2	0.00 → 정상
3	0.14 → 대세포암
4	0.92 → 편평상...
5	0.00 → 선암
6	1.00 → 편평상...
7	0.02 → 대세포암
8	0.00 → 선암
9	0.03 → 선암
10	0.89 → 편평상...
11	1.00 → 편평상...

image name	image
1	1.png
10	10.png
11	11.png
12	12.png
13	13.png
14	14.png
15	15.png
16	16.png
17	17.png
18	18.png
19	19.png

Predictions

Show probabilities for

대세포암
선암
정상
편평상피암

	Neural Network
1	1.00 → 정상
2	1.00 → 정상
3	0.00 → 대세포암
4	0.00 → 편평상...
5	0.00 → 선암
6	0.00 → 편평상...
	0.00 → 대세포암
	0.00 → 선암
	0.00 → 선암
	0.00 → 편평상...
	0.00 → 편평상...
	1.00 → 정상
	0.00 → 대세포암

image name	image	size
1	1.png	189850
10	10.png	558292
11	11.png	67590
12	12.png	96406
13	13.png	68782
14	14.png	83445
15	15.png	76990
16	16.png	98122
17	17.png	113701
18	18.png	86571
19	19.png	76259
2	2.png	188779
20	20.png	83528